Molecular Modeling

From Virtual Tools to Real Problems

Thomas F. Kumosinski, EDITOR
U.S. Department of Agriculture

Michael N. Liebman, EDITOR
Amoco Technology Company

Developed from a symposium sponsored
by the Division of Agricultural and Food Chemistry
at the 205th National Meeting
of the American Chemical Society,
Denver, Colorado,
March 28–April 2, 1993

Molecular Dynamics of Salt Interactions with Peptides, Fibrous Proteins, and Casein

Thomas F. Kumosinski and Joseph J. Unruh

Eastern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, 600 East Mermaid Lane, Philadelphia, PA 19118

Controversy exists concerning the molecular basis for salt induced solubility (salting-in) of proteins. Specifically, the mechanism for the anionic or cationic interactions with either backbone or side chain groups on the protein is still speculative. We initiated molecular dynamics, MD, calculations to clarify this dilemma. MD results for oxytocin in $CaCl_2$ indicate anions H-bond to the peptide N-H backbone causing a conformational change (from turn and extended structure to loop); which is in agreement with the FTIR spectroscopy of oxytoxin. MD calculations on predicted and energy minimized structures of a tropocollagen template molecule in aqueous $CaCl_2$ suggest that Ca^{2+} or Cl^- bind to the tropocollagen N-H backbone bonds, but unlike oxytocin stabilizes the structure. Finally, MD calculations on the N-terminal half of native and dephosphorylated α_{s1}-casein A showed ion binding and hydration energies in agreement with experimental data.

Biotechnology holds the promise of new product development through the use of new designer-type materials possessing tailor-made functionalities obtained via genetic engineering of proteins and/or creation of new protein functionality by controlling co-solutes (*15*). However, the problem of developing quantitative measures for protein structure-function relationships still remains unresolved. Without knowledge of these relationships, new material developments are of limited value due to the low probability of success.

One important functionality is the salt-induced resolubilization of proteins in solutions of mono and divalent salts at concentrations in excess of 0.5 M. Over the decades, many investigators (Steinhardt and Reynolds) (*19*) have performed experimental studies to provide a mechanistic basis for the salt-induced resolubilization of proteins. In a study on the influence of salt to the conformations of proteins, Von Hippel (*23*) defined salts according to their ability to become potential structure formers or structure breakers with respect

to thermal denaturation. Structure formers increase the temperature for protein thermal denaturation whereas structure breakers decrease this temperature. In these traditional studies, only the helix and disordered conformations were considered.

Timasheff's group (*2,21*), on the other hand, has studied the salt-induced resolubilization of proteins as a function of their ability to increase the preferential binding of salt over water in the vicinity of the protein surface. Remarkably, this classical thermodynamic study has provided a quantitative prediction of the resolubilization of protein at high salt concentrations for both mono and divalent salts. In addition, Robinson and Jenks (*17*), through their binding studies of a model peptide compound (acetyl tetraglycine ethyl ester) which contains no formal charge, speculate that salts at high concentrations bind to the high dipole moment regions of the peptide bonds.

In recent years, the emergence of molecular modeling as a technique for refining existing three dimensional molecular structures or building newly predicted structures has yielded a technology with the capability of studying the molecular basis for structure-function relationships (*8,9*). Not only proteins, but also salts and their interactions with proteins may be studied for their potential effect on structure-function relationships.

We now attempt to define and model this simple functionality relationship for oxytocin, tropocollagen and the hydrophilic domain of α_{s1}-casein A, using computer-generated three dimensional structures.

Using molecular modeling techniques (such as energy minimization and molecular dynamics, MD) the protein-salt-water interactions were simulated for each protein in the presence of enough $CaCl_2$ to mimic a greater than 1M solution. This process should establish the most probable protein-salt binding site, but can only mimic the total amount of salt bound to a protein. Whether this controls the salting-in or salting-out of a protein or has any effect on protein solubility can not yet be established.

Theory

Molecular Modeling. All complex structures employed monomer structures previously refined via energy minimization (*10,11*). They were constructed using a docking procedure on an Evans and Sutherland PS390 interactive computer graphics display driven by the Tripos Sybyl (St. Louis, MO) molecular modeling software on a Silicon Graphics 4200 Unix-based computer. The docking procedure allowed for individual manipulation of the orientation of up to four molecular entities relative to one another. The desired orientations could then be frozen in space and merged into one entity for further energy minimization calculations based on molecular force fields i.e. either the Kollman or Tripos generated force fields. The criterion for acceptance of reasonable structures was determined by a combination of experimentally

determined information and the calculation of the lowest energy for that structure.

Force Field Calculation. Studies concerned with the structures and/or energetics of molecules at the atomic level require a detailed knowledge of the potential energy surface (i.e., the potential energy as a function of the atomic coordinates). For systems with a small number of atoms, quantum mechanical methods may be used, but these methods become computationally intractable for larger systems (e.g., most systems of biological interest) because of the large number of atoms that must be considered. For these larger systems, molecular mechanics methods are used. Molecular mechanics is based on the assumption that the true potential energy surface can be approximated with an empirical potential surface consisting of simple analytical functions of the atomic coordinates. The empirical potential energy model treats the atoms as a collection of point masses that are coupled to one another through covalent (bonded) and noncovalent (nonbonded) interactions. The potential energy function (*6,19*) generally has the form:

$$E_{total} = \sum_{bonds} K_r(r-r_{eq})^2 + \sum_{angles} K_0(\theta-\theta_{eq})^2 + \sum_{diherals} \frac{1}{2}K[1+\cos(n\phi-\gamma)] + \sum_{i<j} \frac{B_{ij}}{R_{ij}^{12}} - \frac{A_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \quad (1)$$

The first three terms are due to covalent interactions and represent the difference in energy between the geometry of the actual structure and a geometry in which the bond lengths, bond angles, and dihedral angles all have ideal values. The remaining terms represent nonbonded van der Waals and electrostatic interactions. In Equation 1, r, θ, ϕ, and R_{ij} are variables, determined by the atomic coordinates. All other entities are constant parameters chosen to reproduce experimental observations as closely as possible. Although empirical potential energy functions such as Equation 1 are relatively crude, they have been applied successfully to the study of hydrocarbons, oligonucleotides, peptides and amino acids, as well as systems containing a large number of small molecules such as water. The Tripos force field in Tripos' Sybyl software package uses the above functional equation, and also includes a bump factor which allows atoms to approach each other within a fraction of the van der Waals radius to compensate for H-bonding if so chosen by the user. The parameters used for electrostatic calculations include atomic partial charges (q_i) calculated by the Kollman group (*7,22*) using a united atom approach with only essential hydrogens. All molecular structures were refined with an energy minimization procedure using a conjugate gradient algorithm, in

which the positions of the atoms are adjusted iteratively so as to achieve a minimum potential energy value. Energy minimization calculations were terminated when the energy difference between the current and previous iterations was less than 1 kcal/mol. A nonbonded cutoff of 5 Å (i.e. all non covalently bonded interactions were not calculated for distances greater than 5 Å) was used to save computer time, and is an appropriate value to use for a function that varies with distance. A stabilization energy of at least -10 kcal/mol/residue was achieved for all structures, which is consistent with values obtained for energy minimized structures determined by X-ray crystallography.

Molecular Dynamics. In the previous section we considered only static structures. However, the dynamic motion of molecules in solution contributes to their functionality. The molecular dynamics approach is a method of studying motion and molecular configuration as a function of time (*1*). All atoms in the molecule are assigned a kinetic energy through a velocity term which can be related to the local temperature as well as to the average temperature of the system. These calculations can be performed in vacuum or in the presence of a desired number of solvent molecules such as water. Environmental effects of constant temperature and volume (using a periodic boundary condition to confine the calculation within a prescribed volume), can also be handled in the calculations (*18*). For these calculations a force field describing the potential energy is combined with Newton's second law of motion

$$F_i = m_i a_i(t) = m_i \frac{dv_i(t)}{dt} = m_i \frac{d^2 x_i(t)}{dt^2} = -\nabla_i E \qquad (2)$$

where F_1 is the force on atom i which has mass (m_1), velocity (v_1), acceleration (a_i), and position (x_i). ∇_i is the gradient or the derivative with respect to position, t is the time displacement and E is the potential energy of the molecule described by the chosen force field. Equation (*2*) is integrated at various time intervals for the desired molecule using the chosen force field via a prescribed numerical integration method. The time interval chosen must be small in comparison with the period associated with highest frequency of motion within the molecule. This is usually stretching of a bond associated with a hydrogen atom, i.e. one femtosecond. Numerical integration of Equation (*2*) over one femtosecond intervals to 100 psec for a protein molecule of 2000 atoms or more necessitates a fast computer with a large memory capacity. The results of these calculations can mimic the motions of molecules in solution and also yield time dependent geometric parameters. For example the distance from the center of moment for a set of atoms may be related to correlation times derived from NMR, EPR and fluorescence experiments.

Avian Pancreatic Polypeptide

Three dimensional X-ray crystallographic and circular dichromatic (CD) data exist for avian pancreatic polypeptide (1PPT) (Brookhaven Protein Data Bank, 23). Therefore, 1PPT was chosen to ascertain whether the usual molecular force field and dynamics parameters need to be modified to mimic experimental physical-chemical solution studies. The three dimensional structure of 1PPT consists of 45 residues containing a large α-helix, two gentle turns and a proline helix motif. The two helices are anti-parallel with hydrophobic groups existing within their interface. The molecular force field model must match the X-ray structure. The amount of helix in the model agrees with the X-ray crystallographic structure. The CD results of Noelken et al. (*16*) indicate a large amount of helix conformation (i.e., over 80%) under both dimeric and monomeric conditions. This data agrees with the amount of helix calculated using the X-ray crystallographic structure. Thus, the molecular dynamics (MD) calculations in water must maintain the integrity of both helix conformations to mimic the experimental CD results.

Figure 1 shows the ribbon backbone structure at 50 psec (light ribbon) in the presence of water, with a dielectric constant of unity and a cutoff of 8Å for all non covalently bonded interactions, versus the three dimensional structure (dark ribbon). As can easily be seen, at 50 psec the proline helix structure starts to unwind from the C-terminal end. Such a destabilization in the proline helix is unacceptable when compared with the solution structural CD results. In addition, the time dependence of the radius of gyration was not constant. The use of the large non-interaction cutoff distance of 8 Å with a dielectric constant of unity assumes that this system maintains a uniform dielectric region. While this assumption is appropriate for water, it is well known that proteins do not possess a uniform dielectric constant throughout their structure (*20*). Thus, this approximation predicts a destabilized secondary structure as seen in Figure 1.

Figure 2 shows the same comparison at 50 psec but with a dielectric constant which varies with distance and a cutoff of 5Å for all non-bonded interactions. Here, all helix and turn structures are preserved (50 psec as light ribbon and original X-ray structure as dark ribbon). This holds true even when a Tripos force field (either with all the hydrogen atoms or with only the essential hydrogen bonding atoms) is used instead of the Kollman force field. In addition, the radius of gyration and root-mean-square fluctuation of all atoms becomes constant at 20-25 psec for all calculations. Also, using a larger cutoff value for the non-bonded interactions with additional water molecules did not change the structure of 1PPT. Hence, for the sake of computer time and speed, the lower cutoff value of 5Å was employed for all calculations. Additionally, the appropriate number of water molecules to maintain a density of unity and a periodic boundary was included in all calculations.

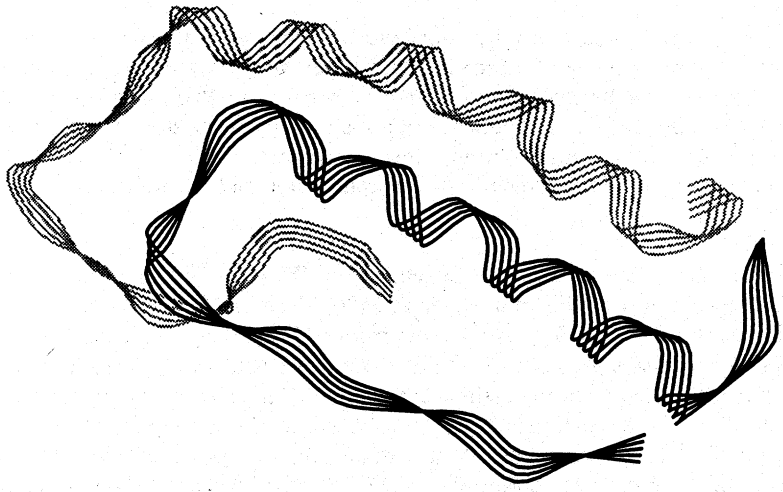

Figure 1. Ribboned backbone dynamic structure of avian pancreatic polypeptide. Crystallographic structure: dark ribbon. Molecular dynamic structure at 50 psec and 300 °K in water with dielectric constant of unity: light ribbon.

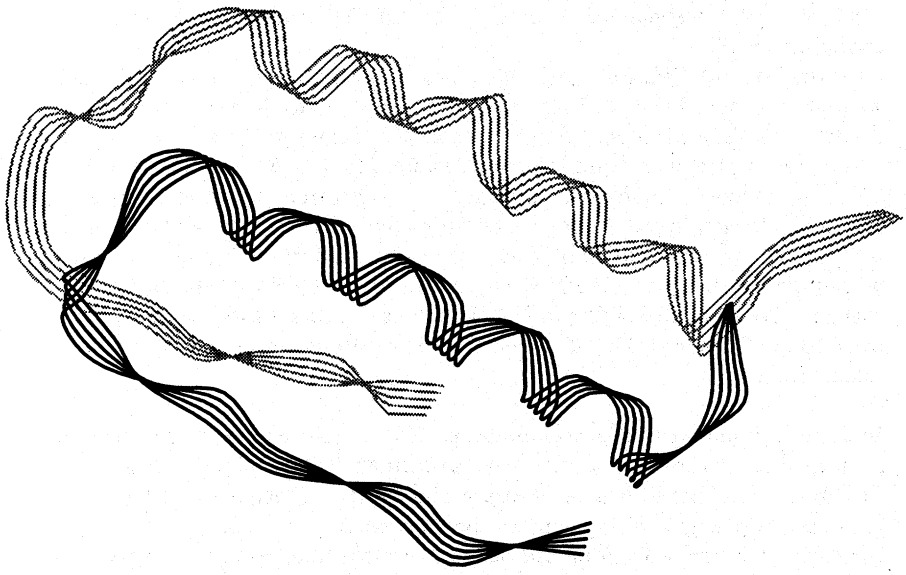

Figure 2. Ribboned backbone dynamic structure of avian pancreatic polypeptide. Crystallographic structure: dark ribbon. Molecular dynamic structure at 50 psec and 300 °K in water and utilizing a dielectric constant that varies with distance i.e. D=R: light ribbon.

Oxytocin

To test the molecular dynamics calculations for modeling the salt induced conformational change of a protein, we will use the X-ray crystallographic structure and results from the Fourier Transform Infra Red Spectroscopy of oxytocin. Oxytocin is a peptide of nine residues with one disulfide bond and two β-turns. It is available in pure form as an acetate salt. The biological role of oxytocin is to induce smooth muscle contraction and to initiate lactation in all mammals.

Fourier Transform Infra Red Spectroscopy (FTIR). Figure 3A shows the FTIR spectrum of oxytocin in water at 25°C. The amide II envelope, which ranges from 1520 to 1580 cm^{-1} results from N-H and N-C deformations of the backbone peptides groups. The amide I envelope which ranges from 1620 to 1700 cm^{-1} results from the carbonyl stretching of the backbone peptide groups. Both the amide I and II envelopes are sensitive to the conformational state of the peptide bonds in proteins. In this figure, three large absorbances at 1676, 1615 and 1556 cm^{-1} are observed. A smaller absorbance at 1510 cm^{-1} caused by the tyrosine side chain O-H deformation also appears. Analysis of the envelopes into component Gaussian bands was not considered necessary for the interpretation of these results. The 1676 cm^{-1} envelope has been assigned by Krim and Bandakar (*8*) as a turn conformation. The 1610 cm^{-1} envelope arises from the carbonyl group stretching of the acetate anion within the sample, while the 1555 cm^{-1} is assigned to the amide II peptide deformation for a turn conformation (*8*).

When 0.05M $CaCl_2$ is added, the amide II absorptivity is almost eliminated and only one absorbance at 1645 cm^{-1} in the amide I envelope is observed (Figure 3B). The 1645 cm^{-1} frequency band has been classified by Byler and Susi (*3*) as an irregular or disordered structure. Hence, when 0.05M $CaCl_2$ is added to oxytocin a conformation change is induced from a turn to an irregular or disordered non-hydrogen bonded conformation. In addition, since the amide II absorptivity is virtually eliminated, the conformational change must be caused by salt binding of at least the N-H backbone group. Such a conformational change caused by salt binding to N-H groups is in accord with the results of Robinson and Jencks (*17*) and should be possible to model using MD calculations.

Molecular Dynamics (MD) Simulations. To simulate the above salt-induced conformational changes in oxytocin we performed MD calculations until equilibrium was established (as measured by the time dependence of the geometric parameters of oxytocin) in the absence and presence of $CaCl_2$ atoms. Here the 3d structure from the Brookhaven Protein Data Bank was in actuality β-mercaptoproprionate oxytocin (wet form) which was modified to cysteine for compatibility with the experimental protein results. The modified structure was

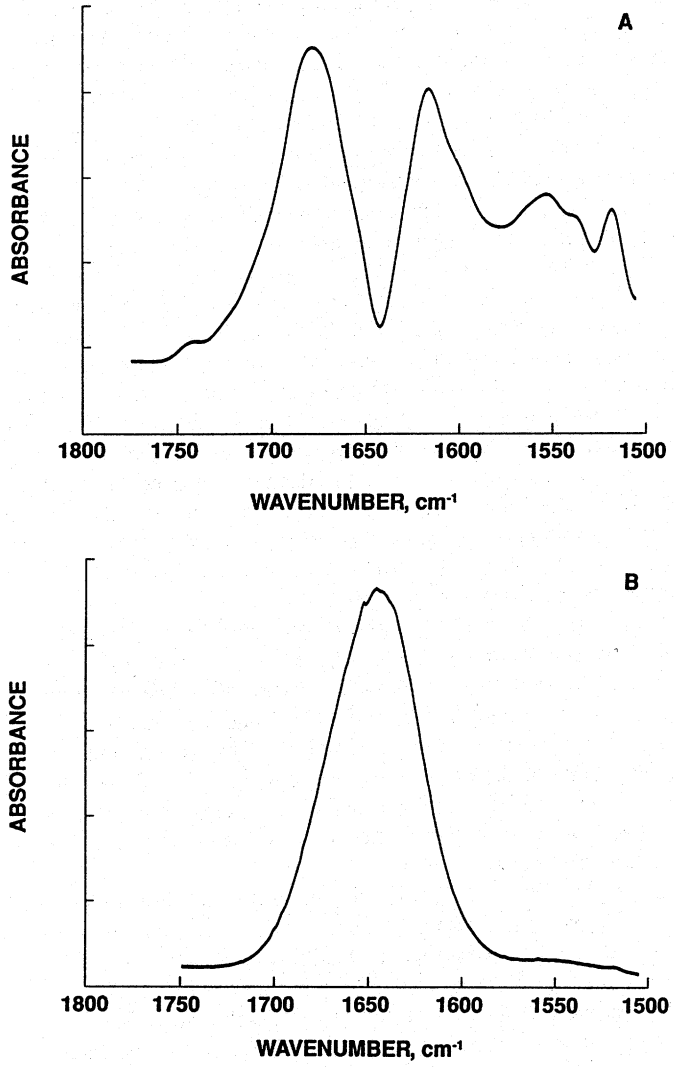

Figure 3. Fourier transform infra red spectra of oxytocin A: in water; and B: in 0.05M $CaCl_2$.

then energy minimized before performing MD calculations. It should be noted that no change in the two β-turns (of the original Brookhaven structure) resulted from this theoretical modification. In addition, in order to mimic salt binding to the backbone N-H group, we allowed the chlorine atoms to become hydrogen bond acceptors. The Tripos force field was used since it contains a parameter which allows a hydrogen atom to come within a specific fraction of the van der Waals radius of the chlorine atom. This fraction is called a bump factor. For these calculations, as well as all other MD calculations in this paper, a default value of 0.7 was maintained for the bump factor. The number of water molecules for all MD calculations was 400. When present in any MD calculations, $CaCl_2$ atoms were docked into the models as Ca and Cl atoms with their van der Wall radii and charges adjusted to mimic Ca^{2+} and Cl^- ions.

Figure 4A shows the variation of oxytocin with time during the MD calculation at 300 °K. As can be seen, the system become equilibrated within 70 psec by the invariancy with time of the radius of gyration, r, and root-mean-square fluctuations, a, of all atoms of oxytocin. However, throughout this simulation the tow turn conformation of oxytocin remains constant as exhibited by the constancy of the internal hydrogen bonds shown in Figure 4B. Figure 4B shows the dynamic structures of oxytocin at 70 psec in a wireframe structure with water (water not shown) and internal hydrogen bonds represented by dashed lines. Here, the backbone structure remains essentially the same as the backbone of the X-ray crystallographic structure. Only the orientation of the side chains have changed due to the MD calculations. Also, shown in Figure 4B is the ribbon backbone for the oxytocin molecule.

Next, 11 Ca^{2+} and 22 Cl^{-1} ions were docked into the equilibrated oxytocin-water system of Figure 4B and energy minimized. The resulting oxytocin with a Cl^{-1} Ca^{2+} and Cl^{-1} distribution is shown in Figure 4C. The oxytocin is represented by a ribboned backbone and the Ca^{2+} and Cl^{-1} ions by balls with radii 0.15 times that of their van der Waals radii. The smaller radii simplified the cumbersome nature of the system. It also should be noted that the Ca^{2+} and Cl^{-1} ions display a different distribution than seen in Figure 4D after MD calculations. The latter distribution is closer to the ribbon oxytocin molecule and generally more compact.

In Figure 4D, 11 Ca^{2+} and 22 Cl^{-1} ions were docked into the oxytocin-water system with a periodic boundary, energy minimized and subjected to MD calculations at 300 °K for 70 psec where equilibrium was easily established. The results show that the chloride ions have hydrogen bonded to the NH backbone atoms as well as the OH group of the tyrosine side chain. Since only a bump factor of 0.7 was utilized, electrostatics appear to be the interaction responsible for this phenomenon. The chloride ion also binds through electrostatics to the Ca^{2+} ion which in turn binds to another Cl^{-1} ion until a network of salt bridges is created with the end Cl^{-1} hydrogen bonded to the backbone N-H groups. In fact all N-H groups are hydrogen bonded to a salt bridge network which destroys both turn conformations and creates a loop-type or irregular structure. This change in conformation is illustrated in Figure 4E,

where the turn conformation is shown as a backbone ribboned structure and the loop conformation is a wireframe representation with Cl^{-1} bonded to N-H groups as dashed lines. It is apparent that the inertia of the $CaCl_2$ salt-bridge network changed the $-\phi$, ψ angles of the oxytocin to create a more open and disordered structure. It is also surprising that the water molecules did not destroy the salt-bridge network since the Cl^{-1} ions could have hydrogen bonded to the water molecules. The reason for this dilemma may arise from the difference in inertia between a water molecule, which has fast motion about a substantial dipole movement, and the Cl^{-1} ion which is spherically symmetric thus preferring to be associated with a group possessing a lower kinetic energy. It also should be noted that MD calculations of the oxytocin and $CaCl_2$ system when water was eliminated yielded the same results.

The above represents experimental and theoretical studies which remarkably agree with one another. To date, the authors are now aware of any other study that correlates MD calculations with FTIR results. However, for the remainder of this paper only MD calculations will be utilized, since no FTIR observations of $CaCl_2$ interaction with fibrous proteins or the hydrophilic half of α_{s1}-casein A exist.

Tropocollagen

For a fibrous protein model we use the tropocollagen structure. Since no X-ray crystallographic structure for tropocollagen exists in the Brookhaven Protein Data Banks, it was necessary to construct such a structure from ϕ, ψ angles reported in the literature. Here, we utilized the reported ϕ, ψ angles of Miller and Scheraga (*14*), who also constructed a lowest energy tropocollagen with closest packing. A template repetitive sequence of gly-ala-Hpro for a total of 180 residues per strand was constructed using appropriate ϕ, ψ angles. Three such strands were docked together in a closest packing manner with the appropriate translation adjustments to create a super helix motif with a pitch of 32 to 33 residues. This structure was energy minimized and the resulting structure containing the super helix is shown in Color Plate 21. In Color Plate 21 the three chains are colored magenta, green and red-orange respectively, and the atoms are represented by spacefilling for easy perception of the super helix motif. While a gly-ala-Hpro template is utilized in this tropocollagen structure a gly-pro-Hpro template could also be so constructed and would yield the same super-helix structure.

Figure 5A shows the intermolecular hydrogen bonding network as dashed lines holding the three individual strands in a stable tropocollagen structure. it is apparent that any disruption of these intermolecular hydrogen bonds between backbone C-O and NH groups would cause dissociation of the three stranded super helix structure.

MD calculations were performed on a smaller portion of the tropocollagen structure (i.e. 45 residues per helical strand for a total of 135 residues) to

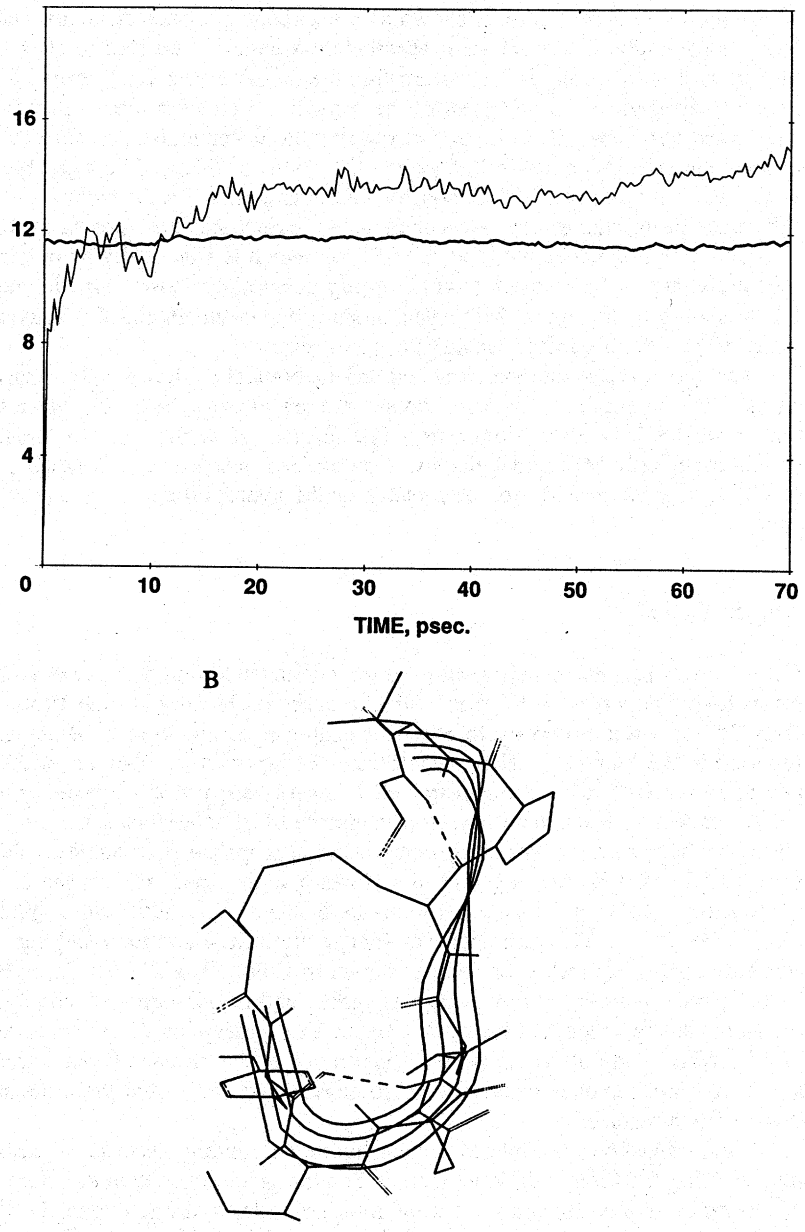

Figure 4. Molecular dynamics of oxytocin at 300 °K. A: Time dependency of geometric parameters of oxytocin: solid line, radius of gyration of all oxytocin atoms, r.; double line: root-mean-square fluctuations of all atoms, a, of oxytocin. B: Wireframe structure with dashed lines as internal hydrogen bonds and ribboned backbone of oxytocin in water at 70 psec.

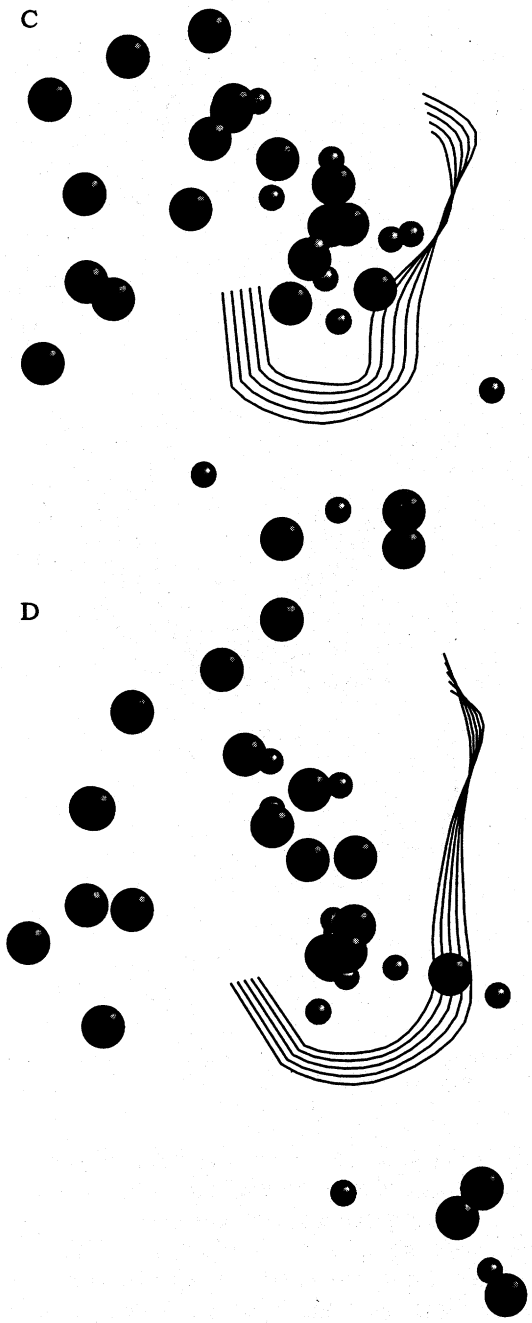

Figure 4. Continued. C: Ball pattern of fluctuations of $CaCl_2$ atoms before molecular dynamics at 0 psec (Ball at 0.15 van der Waals radius). Oxytocin exhibited as a ribboned backbone structure. D: Same as C but after dynamics at 70 psec. *Continued on next page.*

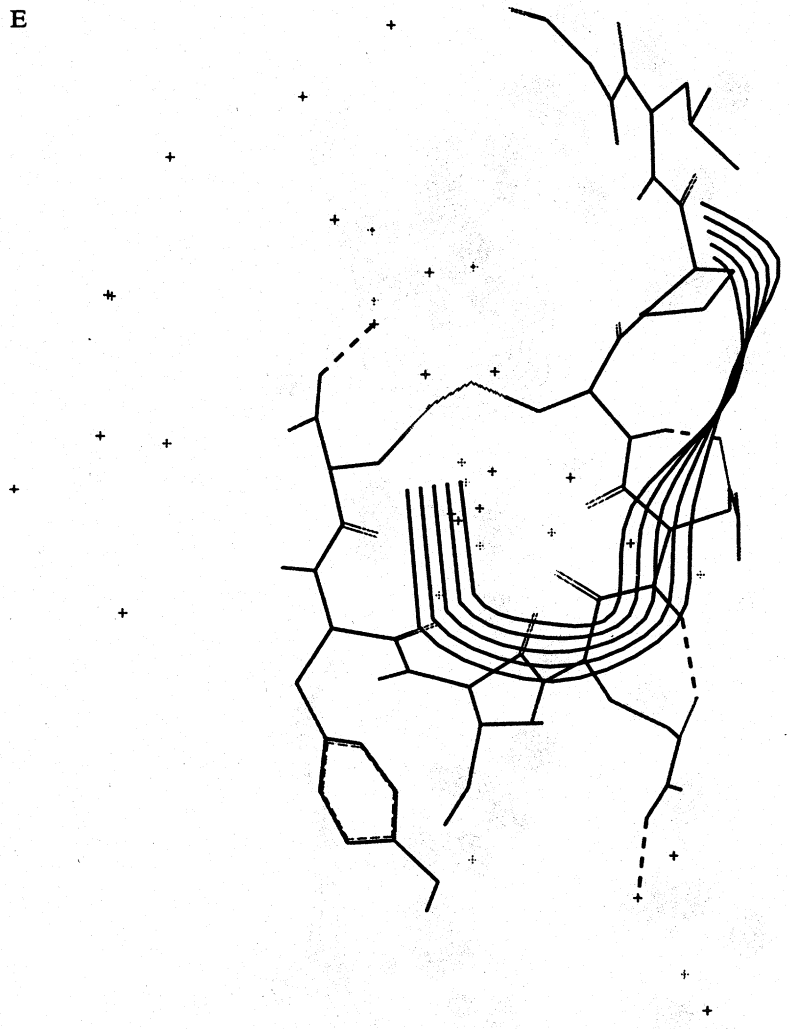

Figure 4. Continued. E: Ribboned backbone of oxytocin at 0 psec wireframe structure with dotted H bond lines at 70 psec.

decrease the time length of the calculation. Four hundred water molecules were added to insure a density of one, along with a non-bonded cutoff of 5Å and a period boundary condition. This small piece would insure at least one turn of the super-helix of the tropocollagen structure. After 50 psec at 300° K, equilibrium was well established and no disruption in the tropocollagen structure (ribboned structure of Figure 5B) was observed. Thus, this small portion of the tropocollagen structure is dynamically stable at 300 °K and 50 psec. Figure 5B also shows the added Ca^{2+} and Cl^{-1} ions after docking into equilibrated tropocollagen and water system.

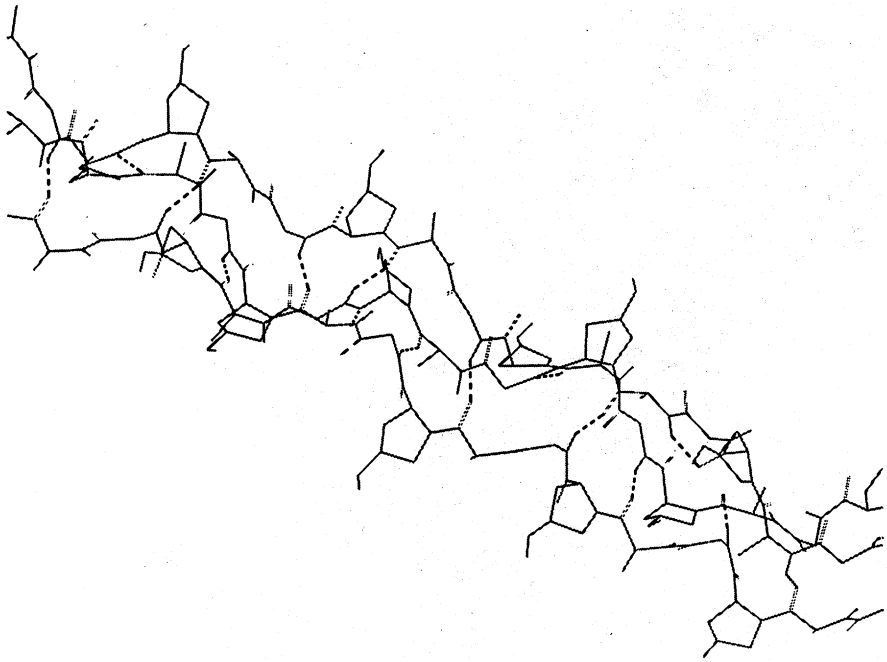

Figure 5. Three-dimensional structure of Tropocollagen. A: Wireframe structure showing dashed lines as hydrogen bonding. *Continued on next page.*

Addition of $CaCl_2$ to mimic a 3M solution (i.e. 22 Ca^{2+} and 44 Cl^{-1} atoms) was then achieved using the docking and merge procedure in Sybyl. The system was energy minimized and submitted to MD calculations at 300 °K for 50 psec where equilibrium was well established. Figure 5C shows a ribboned backbone of tropocollagen with the Ca^{2+} and Cl^{-1} represented as balls (as in Figure 5B) after 50 psec with added Ca^{2+} and Cl^{-1} ions. Here, it is seen that no disruption in the tropocollagen's structure occurs throughout any of the dynamic calculations. The Cl^{-1} of a stable salt-bridge network of $CaCl_2$ hydrogen bond to the solvent-exposed backbone N-H groups. However, unlike in the oxytocin case, the $CaCl_2$ salt bridge network tends to stabilize the tropocollagen structure and may even cause a less dynamically flexible structure. It also should be noted that the distribution of the Ca^{2+} and Cl^{-1} are more compact and closer to the ribboned tropocollagen structure after MD calculations for 50 psec (see Figure 5C) than prior to MD calculations (see Figure 5B). The salt-bridge network can easily be seen in Figure 5C where the Ca^{2+} and Cl^{-1} ions are represented as dotted van der Waals surfaces and tropocollagen as a ribboned backbone structure as in Figure 5B.

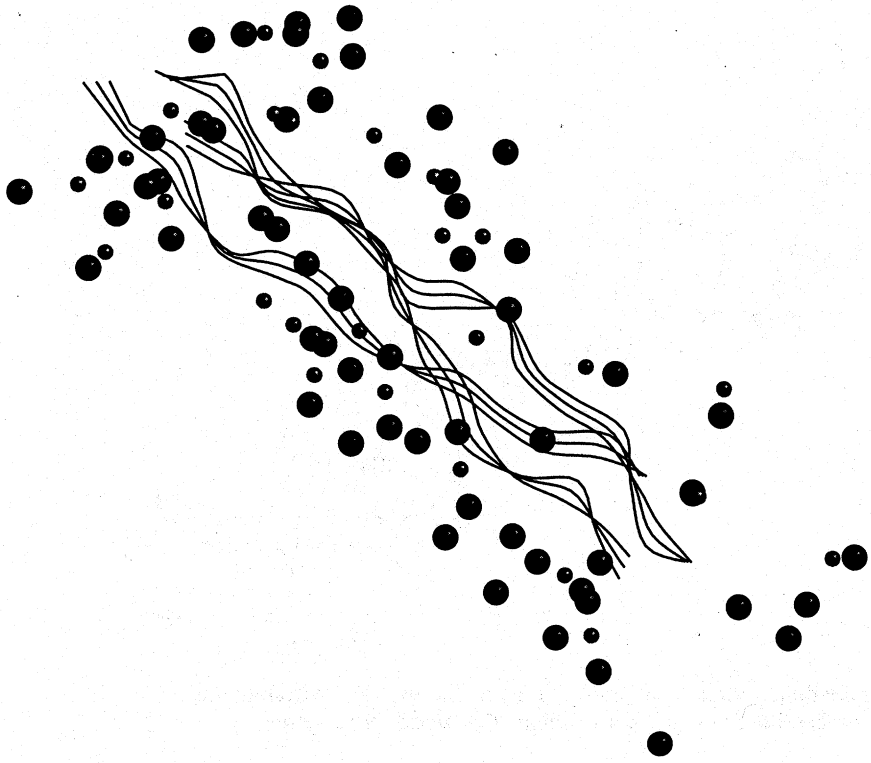

Figure 5. Continued. Three-dimensional structure of Tropocollagen. B: Dynamic structure at 300 ° K and 0 psec for a 135 residues tropocollagen ribboned backbone structure in water with added $CaCl_2$ representing balls and no water shown.

It should be stressed that, although MD calculations show a large salt-bridge network in all structures, the calculation does not take into account the translational diffusion of the salt ions. In reality such a large network most probably would not exist due to the translational diffusion of the salt.

Hydrophilic domain of α_{s1}-casein A

Molecular Dynamics of $CaCl_2$ and $MgCl_2$ in Water. In this section we shall attempt to form a structural basis for the thermodynamics of the salting-in process which was characterized using thermodynamic linkage and non-linear regression analysis of the calcium-induced solubility profiles of α_{s1}-casein (*6,13,13*). We have chosen the hydrophilic N-terminal domain of the monomeric α_{s1}-casein A structure to model this solubility profile. This model was built from the α_{s1}-casein B model (*12,13*) by excising residues 14 through 26 and then deleting residues 100 to 199 of α_{s1}-B. The resulting hydrophilic domain, i.e., residues 1 through 99 of the α_{s1}-casein B minus 14 to 26 (a total of 86 residues), was then energy minimized. Eleven hundred water molecules

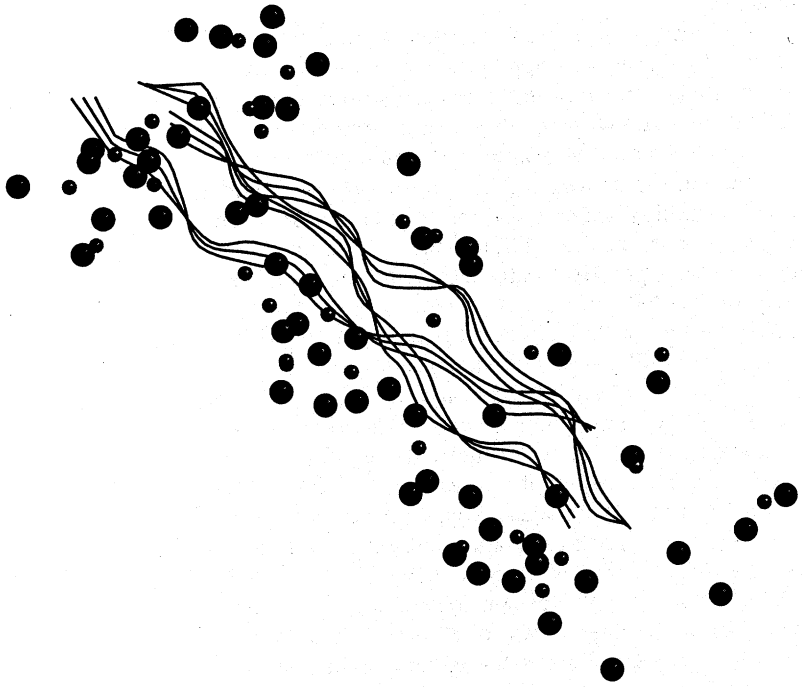

Figure 5. Continued. Three-dimensional structure of Tropocollagen. C: Dynamic structure of B at 300 ° K and 50 psec.

were added and the resulting structure in solution was again energy minimized with a cutoff for all non-bonded interactions of 5 Å, while maintaining a periodic boundary condition. The resulting structure with water was subjected to molecular dynamics (MD) calculation for 20 psec above equilibrium conditions which were determined by the stabilization of potential energy, radius of gyration of the protein backbone, root-mean-square fluctuations of the backbone atoms, and change in second moment. Such an equilibrated dynamic structure should approximate the structure, energetics and dynamic motion of this protein domain in solution.

To mimic the salt binding mechanism, 22 molecules of calcium or magnesium and 44 molecules of chloride with appropriate ionic charges were added to the above system in a pseudo random fashion. The system was energy minimized and subject to MD for a full 40 psec. Equilibrium was easily established once again at 15 to 20 psec. The resulting structure for the native half of α_{s1}-A in $CaCl_2$ is shown in Figure 6 A and B. The amount of salt added was chosen to comply with a condition which would result in saturation of the Ca^{2+} binding sites. Using a literature value of 380 M^{-1} for K_A (*6,9*) with 8 binding sites, 22 molecules of $CaCl_2$ per 1100 water molecules per molecule of protein is equivalent to greater than 99% occupancy of these calcium binding sites and greater than 80% occupancy with 8 additional putative salting-in sites derived using k_2 from Farrell, et al. (*6*). A similar structure for the dephosphorylated half of α_{s1}-A was also studied and is shown in Figure 7. In total, seven MD calculations were performed on the hydrophilic domain (H) of α_{s1}-A (residues 1-99) in the presence of 1100 water molecules to 40 psec: two in the absence of salt for native (H) and dephosphorylated (HO-P); two in the presence of $CaCl_2$ for H and HO-P; one with added $MgCl_2$ for H; and one each for $MgCl_2$ and $CaCl_2$ with no protein. Each calculation utilized a cutoff of 5 Å for non-bonded interactions, a Tripos force field, and a "bump" factor of 0.7 for simulating hydrogen bond formation. MD calculations require a running time on the Silicon Graphics Unix computer system of at least 3 days. It should be noted at this time that the energetics and geometric parameters estimated by the MD calculation reflect the total salt binding to the protein, i.e., the sum of the free energies of salt binding for the protein salting-out, as well as the salting-in process. More detailed analysis of these MD calculations, which are beyond the scope of this paper, must be performed for separation of the protein precipitation and resolubilization processes.

The average results of several calculated geometric parameters with their corresponding errors are presented in Table 1. The subscript 1 denotes the salt component i.e. both Ca^{2+} and Cl^{-1}, while the subscript 2 describes the protein. R is the calculated dynamic radius of gyration, x is the average center of mass of the component atoms and a is the root-mean-square fluctuation of component atoms from the center of mass. Here, a can be thought of as the dynamic Stokes radius of the chosen ions, and x is the spherical center of mass.

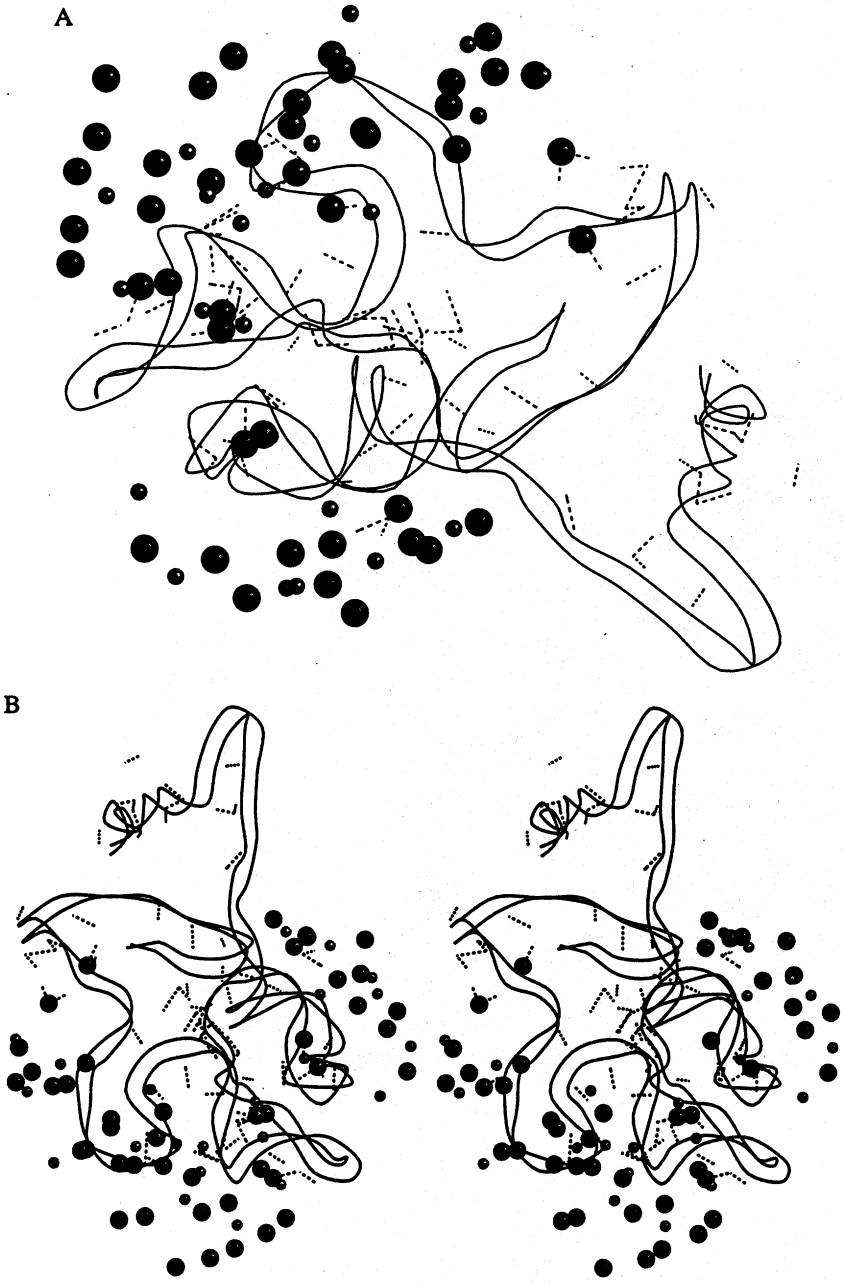

Figure 6. A: Backbone ribbon structure of the hydrophilic half of α_{s1}-casein A (residue 1 through 86) after molecular dynamics at 40 psec with 22 molecules of $CaCl_2$ in the presence of 1100 water molecules. Ca and Cl atoms are shown as ball models of radii equal to 0.15 times their known van der Waals radius. Dashed lines represent hydrogen bond formation. B: Stereo view (relaxed view) of A.

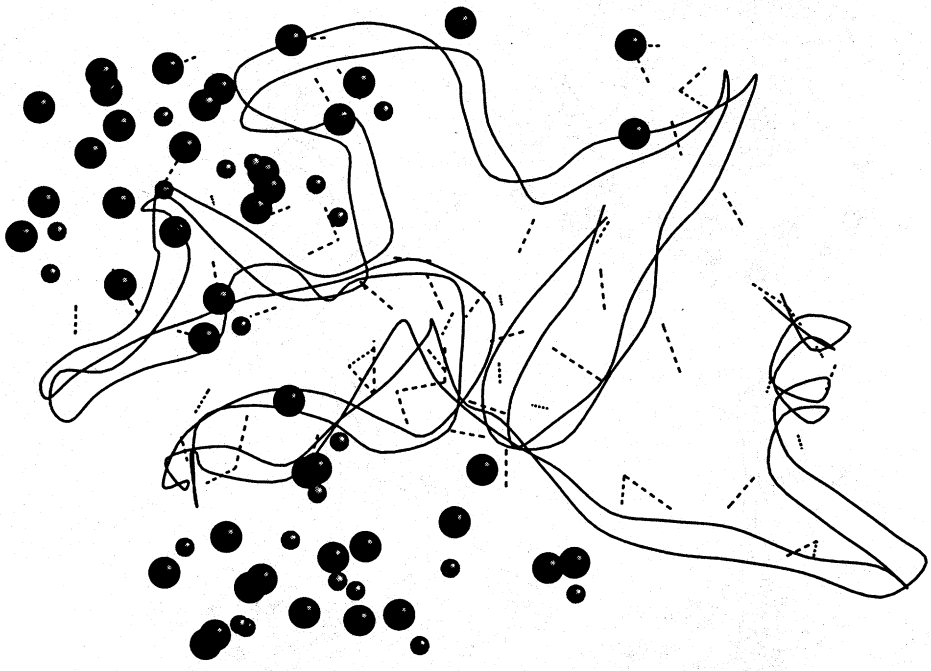

Figure 7. Same as in Figure 6A but for dephosphorylated α_{s1}-A. All serine phosphates mutated to serine with appropriate partial charges.

Table I. Molecular Dynamics of Hydrophilic Half (H) of α_{s1}-casein A in Water with Salt: Geometric Parameters

Protein	Salt	a_1, Å^2	x_1, Å	R_2, Å	a_2, Å^2	x_2, Å
H	—			14.6 ± 0.2	8.2 ± 0.5	0.33 ± 0.04
HO-P	—	—	—	14.2 ± 0.04	7.0 ± 0.2	0.33 ± 0.03
H	$CaCl_2$	11.5 ± 0.7	0.8 ± 0.10	14.6 ± 0.2	8.7 ± 0.4	0.38 ± 0.04
HO-P	$CaCl_2$	9.38 ± 0.67	0.69 ± 0.06	15.5 ± 0.2	10.1 ± 0.4	0.51 ± 0.03
H	$MgCl_2$	7.8 ± 0.6	0.67 ± 0.07	14.6 ± 0.2	7.7 ± 0.5	0.37 ± 0.04
—	$CaCl_2$	9.1 ± 0.9	2.2 ± 0.1	—	—	—
—	$MgCl_2$	6.8 ± 1.1	1.5 ± 0.4	—	—	—

R is radius of gyration
a is the RMS fluctuation of all atoms from center of mass; a dynamic Stokes radius.
x is the calculated spherical center of mass.
subscript 1 denotes salt while 2 denotes protein atoms.

Using Table 1 we can attempt to describe the distribution of the salt ions at the end of the MD calculations. To do so, we inspect the a and x values in Table 1 for both $CaCl_2$ and $MgCl_2$ alone and in the presence of both the hydrophilic half of α_{s1}-A (H) and its dephosphorylated form HO-P. No significant changes are seen in the a_1 values, since the mass of the proteins is far greater than that of the salts. The x_1 values are the best descriptors of the binding of ions to the proteins in these MD calculations. In all cases the x_1 values have decreased in the presence of either protein component. Such a decrease in the average spherical center of mass of the ions is a clear indication of protein-salt interactions, since the center of mass for salt ions alone would be larger as they move randomly about. However, x_1 would be smaller for salt bound to protein where movement is restricted. A difference between x_1 for the native H and the O-P form exists. It can easily be observed (Figures 6 and 7) that the salt ions associated with the H and HO-P forms have a different respective overall distribution. This difference needs to be verified but it is in agreement with the decreased x_1 in the presence of the protein fragments. It is assumed that the x_1 of the $MgCl_2$ was smaller due to the smaller value of the van der Waals radius for Mg^{2+} (0.66 Å) vs. Ca^{2+} (0.99 Å).

To observe the effect of salt binding on the dynamic structure of the hydrophilic half of the protein, we have calculated for the protein (in the presence or absence of salt ions) its radius of gyration R_2, the a_2 and the x_2 values. These values are presented as columns 5, 6 and 7 of Table 1. Virtually no changes within the calculated error are observed for the R_2, a_2 or x_2 values for the native H form either in the absence or presence of $CaCl_2$ or $MgCl_2$. The HO-P form is not dramatically different either. However large increases in these descriptors are observed in the dephosphorylated HO-P form when $CaCl_2$ is added in the MD calculations. This change may reflect a general expansion of the HO-P structure when $CaCl_2$ is added which can also be observed by inspection of Figure 8A. Here the two structures are compared by representations of the protein backbones, with no side chains displayed. The H form is represented by a ribbon trace of the backbone, while the HO-P form is represented by a backbone-atomic stick model. The backbone model is much more expanded than the ribbon model. For easier observation of this conclusion a stereo view is shown in Figure 8B. The reason for this phenomenon is most likely due to the hydrogen bonding of chloride ions to the serine side chains as well as to N-H atoms of the backbone. Such anion-protein hydrogen bonding can impose important dynamic structural changes on the protein component. In the H form these interactions may be "screened-out" by the negatively charged phosphate groups. Whether this interaction causes increased solubility of the protein (i.e. salting-in) cannot be established at this time, because the potential role of the deleted residues is not considered. More MD calculations in conjunction with physical chemical experiments (e.g. FTIR of proteins in solutions) must be performed in the future to test this hypothesis.

To further correlate the binding free energies calculated from the literature for thermodynamic linkage analysis of the salting-out and salting-in

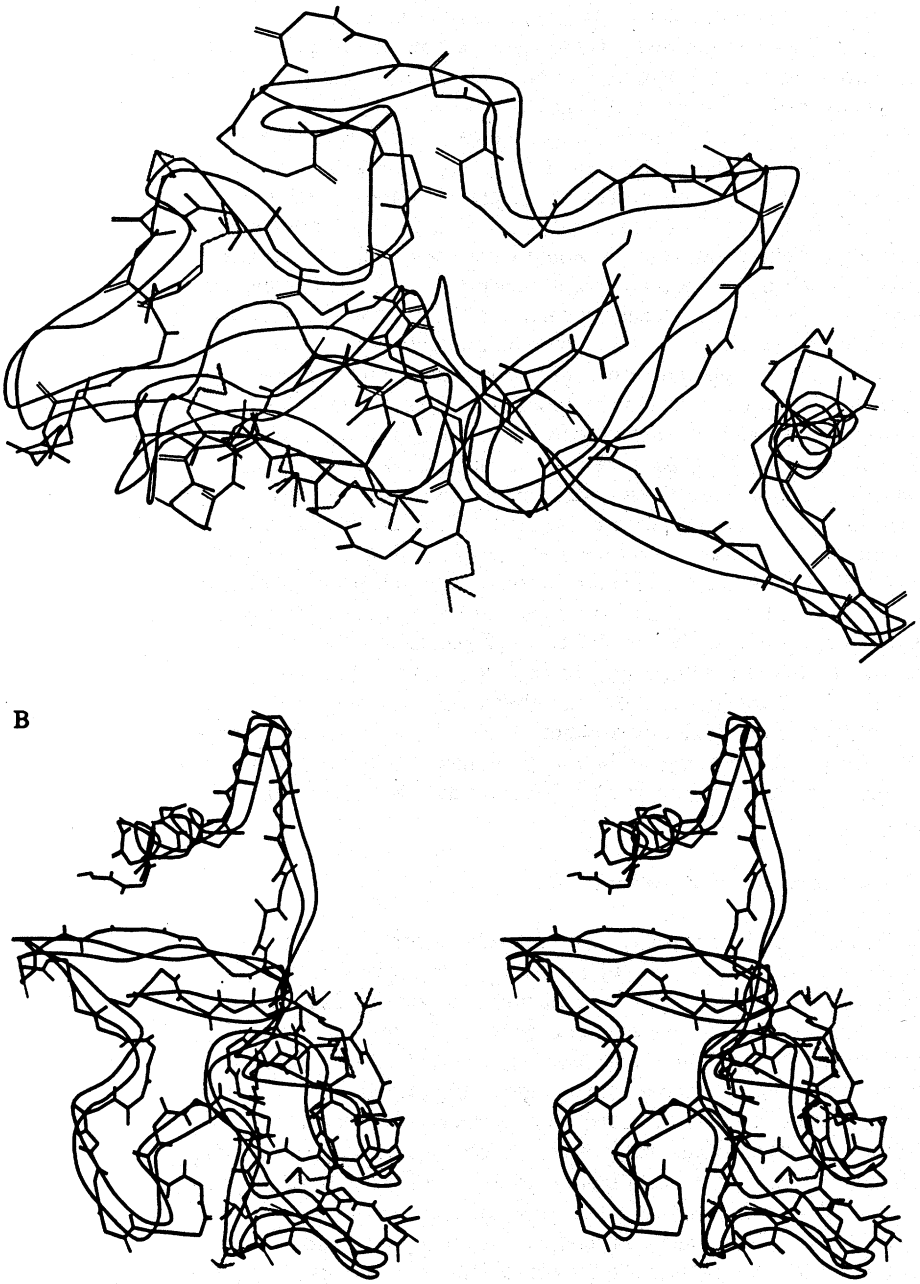

Figure 8. A. Comparison of hydrophilic domain of α_{s1}-A (shown with ribboned backbone) with dephosphorylated α_{s1}-A after molecular dynamics with 22 molecules of Ca and Cl_2 at 40 psec in the presence of 1100 water molecules dashed lines represent hydrogen bond formations. B. Stereo view (relaxed) of A.

experiments, ΔF_B, (*6*), the energetics of the seven MD calculations for the modeled H and HO-P were calculated. These values are presented in Table 2 along with the experimental total binding free energies, ΔF_B calculated from Farrell et al. (*6*), for native α_{s1}-A in $CaCl_2$ and $MgCl_2$ as well as for dephosphorylated α_{s1}-A in $CaCl_2$. In this table, the calculated descriptors of energy are: E_T, for the average total potential energy of the system; E_W, for the water; E_P, for the energy of the protein; E_{PW}, protein-water interaction energy; E_{SW}, for the salt-water interaction energy; and E_{INT}, for the potential energy for the salt-protein interaction. It should be stressed that the energies estimated from MD calculations are the internal energy at constant volume and not the Gibbs or the Helmholtz free energy derived from binding experiments. However using the approximation that the potential interaction energy is equal to the ΔF_B, a qualitative correlation between these two parameters can be utilized to describe variations in protein components or type of salt used. Inspection of columns 4 and 5 of Table 2 shows that for native α_{s1}-A and HO-P in $CaCl_2$ and native A in $MgCl_2$, the experimental changes of ΔF_B correlate well with changes in ΔE_{INT} when these environmental factors are varied; the absolute values for the two parameters are, however, significantly different. The large differences in absolute values between ΔF_B and ΔE_{INT} are due to the bump factor of 0.7 used in these calculations to quantitate hydrogen bonding in MD. If this factor were optimized then perhaps better correlations could be achieved. However, for this study, the value of 0.7 was used since it is the default value used by Sybyl for the Tripos force field. Such a small bump factor most likely allows for inordinately high electrostatic energy terms. Future work will tend to optimize this parameter. No conclusions can be made at this time concerning the meaning of the other energetic descriptors of Table 2 but they are presented along with their error for inspection by the reader.

Summary Correlations and Conclusions

From all the above, it can be concluded that molecular modeling techniques such as energy minimization and molecular dynamics can provide a powerful multifaceted approach for defining structure-function relationships such as the salt-protein interactions. Here, the dynamic changes in the protein structure responsible for the salt binding processes can be established using energy minimization and molecular dynamics calculations of the protein in water, in the presence and absence of salt ions. In particular, predictions concerning the type of protein primary structure groups responsible for binding and conformational change which occurs can be utilized to increase the desired protein functionality in a rational way through chemical or genetic modification. (*15*).

The molecular dynamics calculations could not at this present time distinguish between the salting-out and salting-in binding free energy. Only the total binding could be correlated in the MD results. In fact, significant

Table II. Molecular Dynamics of Hydrophilic Half (H) of α_{s1}-casein A in Water with Salt: Energetics and Comparison with Experimental Data

Protein	Salt	$-E_T$	$-E_{int}$	ΔF_B	$-E_{PW}$	$-E_W$	$-E_{SW}$
H	—	20,000 ± 152	—	—	20,000 ± 152	13,000 ± 64	—
HO-P	—	19,000 ± 54	—	—	19,000 ± 54	13,000 ± 101	—
H	$CaCl_2$	31,000 ± 134	8,746 ± 162	33.2	13,000 ± 154	11,000 ± 142	20,000 ± 128
HO-P	$CaCl_2$	28,000 ± 132	4,612 ± 198	14.3	14,000 ± 26	11,000 ± 91	22,000 ± 155
H	$MgCl_2$	34,000 ± 92	11,000 ± 175	36.6	12,000 ± 133	11,000 ± 142	22,000 ± 155
—	$CaCl_2$	27,000 ± 266	—		—	13,000 ± 122	27,000 ± 266
—	$MgCl_2$	29,000 ± 198	—		—	12,000 ± 198	29,000 ± 198

E denotes potential energy in kcal/mole i.e. internal energy at constant volume subscripts:
T denotes total atoms
P denotes protein atoms
W denotes water atoms
S denotes salt atoms and int denotes protein-salt interaction.
ΔF_B denotes the total co-operative salt binding free energy in kcal/mole calculated from Tables 5 and 6 in Ref. 6), for the solubility data for α_{s1}-casein A (NA) and its dephosphorylated form (NAO-P).

differences in absolute values occur when the total salt binding free energy calculated from thermodynamic linkage, ΔF_B of the hydrophilic half of α_{s1}-casein A, is compared with the interaction energy derived from MD, ΔE_{INT} (see Table 2). This discrepancy is most likely due to the bump factor value chosen for the MD calculations; a lower value should be chosen for future work. However, the changes in ΔF_B with the change in salt from Ca to Mg for the native systems as well as the change with calcium between native and the dephosphorylated form show good correlations with ΔE_{INT}. More importantly, the geometric parameters for the proton from MD calculations (i.e. R_2, a_2, and x_2) of the HO-P change dramatically over the H form in the presence of $CaCl_2$. This phenomenon was interpreted as an expansion of the dephosphorylated form with added $CaCl_2$ which is shown in Figure 8A and 8B. It is apparent that this structural change can also be labeled as a conformational change since many internal hydrogen bonds are disrupted. In the future, this phenomenon will be tested experimentally by FTIR.

Literature Cited

1. Andersen, H. C. *J. Chem. Phys.* **1980,** 72:2384-2394.
2. Arakawa, T.; Timasheff, S. N. *Biochem.* **1984,** 23:5912-5923.
3. Byler, D. M. and Susi, H. *Biopolymers.* **1986,** 25:469-487.
4. Eigel, W. N.; Butler, J. E.; Ernstrom, C. A.; Farrell, H. M., Jr.; Harwalkar, V. R.; Jenness, R.; Whitney, R. McL. *J. Dairy Sci.* **1984,** 67:1599-1631.
5. Farrell, H. H., Jr.; Thompson, M. P. The caseins of milk as calcium binding proteins, pp. 117-137. In "Calcium Binding Proteins," M. P. Thompson (Ed). CRC Press Boca Ratan, FL, 1987.
6. Farrell, H. M., Jr.; Kumosinski, T. F.; Pulaski, P.; Thompson, M. P. *Archives Biochem. Biophys.* **1988,** 265:146-158.
7. Kollman, P. A. *Ann. Review Phys. Chem.* **1987,** 38:303-333.
8. Krimm, S. and Bandekor, J. *Adv. Protein Chem.* **1986,** 38:181-364.
9. Kumosinski, T. F.; Farrell, H. M., Jr. *J. Protein. Chem.* **1991,** 10:3-16.
10. Kumosinski, T. F.; Brown, E. M.; Farrell, H. M., Jr. *Trends in Food Sci. Technol.* **1991a,** 2:110-115.
11. Kumosinski, T. F.; Brown, E. M.; Farrell, H. M., Jr. *Trends in Food Sci. Technol.* **1991b,** 2:190-195.
12. Kumosinski, T. F.; Brown, E. M.; Farrell, H. M., Jr. *J. Dairy Sci.* **1993a,** 76:931-945.
13. Kumosinski, T. F.; Brown, E. M.; Farrell, H. M., Jr. *J. Dairy Sci.* **1993b,** 76:2507-2520.
14. Miller, M. H. and Scheraga, H. A. *J. Polymer Sci. Symp.* **1976,** 54:171-200.
15. Noble, R. W.; Waugh, D. F. *J. Amer. Chem. Soc.* **1965,** 87:2236-2257.

16. Noelken, M. E.; Change, P. J. and Kimmel, J. R. *Biochemistry,* **1980,** 19:1838-1843.
17. Robinson, D. R.; Jencks, W. P. *JACS,* **1965,** 87:2470-2479.
18. Sillen, L. G.; Martell, A. E. Stability Constants of Metal Ion Complexes, Special Publication No. 25 of The Chemical Society of London, Alden Press, Oxford, U.K., **1971.**
19. Steinhardt, J. and Reynolds, J. A. In "Multiple Equilibria in Proteins", p. 325. Academic Press, New York, **1969.**
20. Tanford, C. Physical Chemistry of Macromolecules. John Wiley & Sons, New York, **1967.**
21. Timasheff, S. N. and Arakawa, T. In "Protein Structure and Function: A Practical Approach" (Creighton, T. E., Ed.) pp 331-345. IRL Press, Oxford, **1988.**
22. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Cose, D. A. *J. Comput. Chem.* **1986,** 7:230-252.
23. Von Hippel, P. H.; Schleich, T. In Structure and Stability of Biological Macromolecules. S. N. Timasheff and G. D. Fasman, eds. Marcel Dekker, P. 417, New York, NY, **1969.**